

A COMPARISON OF SENSITIVITY ANALYSIS AND ERROR ANALYSIS BASED ON A STREAM ECOSYSTEM MODEL *

R.H. GARDNER and R.V. O'NEILL

Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37830 (U.S.A.)

J.B. MANKIN and J.H. CARNEY

Computer Sciences Division, Union Carbide Corporation, Oak Ridge, TN 37830 (U.S.A.)

(Accepted for publication 6 August 1980)

ABSTRACT

Gardner, R.H., O'Neill, R.V., Mankin, J.B. and Carney, J.H., 1981. A comparison of sensitivity analysis based on a stream ecosystem model. *Ecol. Modelling*, 12: 173–190.

When individual model parameters must be measured in field or laboratory experiments, the provision of feedback information for allocation of research efforts is an important function of modeling. Both sensitivity analysis and Monte Carlo error analysis can be used to determine which parameters require intensified measurement effort. When both methods are applied to a stream ecosystem model, the assumptions of sensitivity analysis are violated if reasonable estimates of measurement errors on parameters are used. Sensitivity analysis estimates a linear relationship between a state variable and a parameter and largely ignores higher order effects.

In the model investigated in this study, higher-order effects dominate prediction error, and the results of sensitivity analysis are misleading. It is suggested that the simple correlation coefficient derived from analysis of Monte Carlo simulations is a more reasonable way to rank model parameters according to their contribution to prediction uncertainty. For the stream model used in this study, halving variance on the four parameters, indicated as most important by sensitivity analysis, reduces prediction errors by only 2–6%. Halving variance on the four, completely different, parameters with the largest simple correlation coefficients reduces prediction errors by 17–31%.

INTRODUCTION

Continued interest in the application of ecological modeling to environmental problems (e.g. Van Winkle, 1977; Jørgensen, 1979) demands increased

* Research supported by the National Science Foundation's Ecosystem Studies Program under Interagency Agreement DEB 77-25781 with the U.S. Department of Energy, Oak Ridge National Laboratory, under contract with the Union Carbide Corporation Environmental Sciences Division Publ. No. 1705.

attention to prediction uncertainty and model limitations. Although uncertainty can result from a number of factors (O'Neill and Gardner, 1979), errors due to parameter estimation are of particular importance to the interaction between the modeler and the field researcher. Typical questions involve the accuracy of field or laboratory measurements required to produce adequate model predictions (O'Neill et al., 1980).

Two approaches to parameter estimation can be distinguished (Gardner and O'Neill, *in press*). In the first approach, one begins with time-series measurements of ecosystem components. Model parameters are fitted to these data by some type of least squares procedure (e.g. O'Neill, 1971; Halfon, 1975; Beck, 1979). Residual errors attributable to parameter estimation are then used to determine the uncertainty associated with model predictions (Beck, *in press*).

A second approach to parameter estimation is emphasized in the present paper. In this case, a complex ecosystem model is derived from population interaction theory and physiological ecology (e.g. Shugart et al., 1974; Innis, 1978). Model size and complexity preclude direct parameter estimation from a single field experiment. Therefore, individual parameters must be determined in independent laboratory or field experiments.

In this second approach, the required accuracy of parameter values can be determined either by sensitivity analysis or by Monte Carlo error analysis. Both analyses attempt to rank order the parameters according to their contribution to overall model error. Sensitivity analysis approaches the problem by taking the partial derivatives of model equations with respect to individual parameters (e.g. Steinhorst et al., 1978). In some cases, which are not considered in this study, the resulting equations may be so complex that direct implementation is impractical. Results are typically stated as the change in a model prediction to be expected from a small error in parameter values. In contrast, error analysis considers each parameter as a random variable. Prediction uncertainty is estimated by Monte Carlo simulation with each parameter selected either from independent distributions (O'Neill, 1973) or from multivariate distributions specified by a covariance matrix (O'Neill et al., 1980). The contribution of each parameter to model uncertainty is determined by statistical analysis of simulation results.

Both sensitivity analysis and error analysis provide feedback to research design. Since both techniques provide the required information, the question arises as to which approach should be applied to a specific problem. The present paper addresses this question, first by contrasting the practicality and assumptions of each approach and showing conditions under which the methods converge to the same result. The methods are then compared directly by applying both to a non-linear model of a stream ecosystem. The question of which approach provides the most relevant feedback information to the ecological field researcher is addressed throughout.

SENSITIVITY ANALYSIS AND ERROR ANALYSIS

Sensitivity analysis has the advantage of being an analytical method (assuming the mathematics permit). Determining the sensitivity of a model output to a specific parameter requires only a single equation which can be numerically solved along with model equations. Error analysis, on the other hand, requires hundreds of model runs before error terms can be estimated. At first sight, then, sensitivity analysis appears to be more efficient and practical.

However, a complex model requires the derivation (and verification) of a large number of sensitivity equations. The four-equation stream model analyzed in this paper requires 44 sensitivity equations. Error analysis, on the other hand, requires only a FORTRAN subroutine for model equations which then is inserted into an existing code for Monte Carlo simulations.

Sensitivity analysis, therefore, utilizes a significant investment of investigator time (in deriving and verifying the sensitivity equations) to save computer time. Error analysis utilizes significantly greater computer time in order to save investigator time. Any cost-benefit comparison between the two methods must, therefore, consider both investigator time and computer expenses.

The relevance of each type of analysis to ecological problems is more important than practicality. To address this question, we must first consider the assumptions implicit in each approach. Because of three implicit assumptions (Gardner and O'Neill, 1981) sensitivity analysis only approximates the contribution of each parameter to overall model error. These assumptions are as follows:

(1) The expected behavior of the model is equal to the behavior of the model using the mean value of each parameter (Argentesi and Olivi, 1976). In past studies, the bias introduced by this assumption has ranged from 1% of total model uncertainty (Gardner et al., 1980a) to 87% (Gardner et al., 1980b).

(2) The error contribution of each parameter can be examined separately. Thus, higher-order effects resulting from simultaneous errors in many parameters (e.g. covariances between parameters) are ignored.

(3) Errors in model predictions can be approximated by examining small perturbations in the parameters. Higher-order effects resulting from large errors are ignored.

Since large variances occur simultaneously in most, if not all, ecological parameters, assumptions (2) and (3) may be seriously violated in many applications. Error analysis makes no a priori assumptions about higher-order effects or the manner in which errors are propagated during a simulation.

Assumptions in error analysis involve the statistical distributions of parameters and their means, variances and covariances. The effects of these assumptions have been examined in previous studies (O'Neill and Gardner, 1979; Gardner and O'Neill, 1980; Gardner et al., 1980b) and will not be considered

further here. The important point is that error analysis is able to relax the assumptions of sensitivity analysis by requiring additional information about model parameters.

A major task of this study, then, is to clarify the conditions under which each technique is appropriate. If assumptions 2 and 3 are significantly violated and the results of sensitivity analysis cannot be trusted, then error analysis should be applied, regardless of the cost of the analysis. If conditions can be identified under which the assumptions of sensitivity analysis are reasonably valid, then the modeler is free to choose between the techniques.

To gain an impression of the conditions under which the two techniques produce the same results, it is useful to examine the simple function, B^4/A^2 , in which both parameters are considered to have a mean value of 1.0. The sensitivity of the function to a change in A is the partial derivative with respect to A , i.e. -2.0 . The equivalent stochastic coefficient is the change in B^4/A^2 divided by the change in A , averaged over the Monte Carlo iterations.

Figure 1 shows the results of calculations in which A was uniformly distributed between limits expressed as a percentage of the mean. Thus, a 20% variability in A means that A was varied uniformly between $A - 0.2A$ and $A + 0.2A$, i.e. $0.8-1.2$. In a sense, this is a test of assumption (3) since we are examining the mean sensitivity coefficient as the variability in A is increased. It is clear that sensitivity analysis is quite robust for this case, since the mean stochastic sensitivity coefficient is equal to -2.0 up to about 30% variability (Fig. 1). As variability in A increases beyond 30%, however, the stochastic coefficient diverges more and more rapidly from -2.0 .

Figure 2 is a test of assumption (2), i.e. where each parameter can be examined separately, since A and B are varied simultaneously, over a range (indicated by the power of 10) shown on the figure. Thus, an error on B of -3 means that B was varied uniformly from $B - (1 \times 10^{-3}) B$ to $B + (1 \times 10^{-3}) B$.

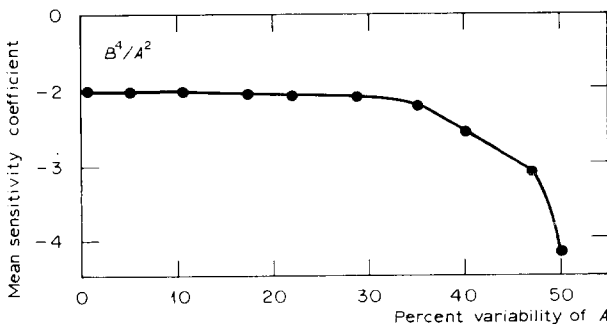


Fig. 1. Mean sensitivity coefficient (deviation in value of the function divided by deviation in parameter A) for B^4/A^2 with different levels of variability in A . Calculations based on 100 Monte Carlo iterations for each data point.

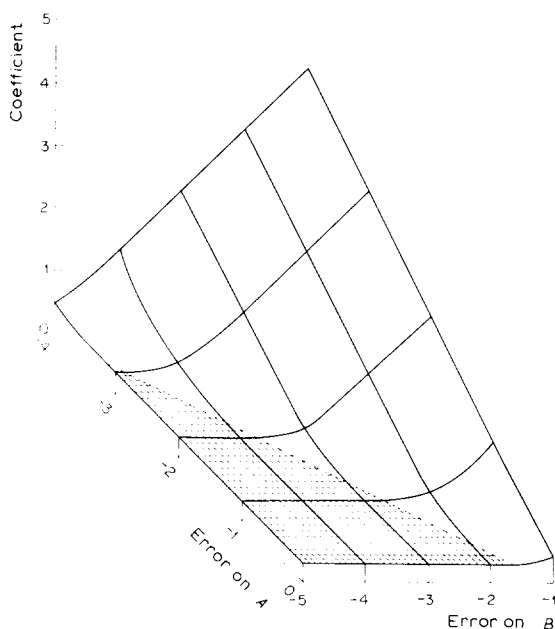


Fig. 2. Mean sensitivity coefficient for B^4/A^2 as a function of variability in B and A . The sensitivity coefficient is the deviation in B^4/A^2 divided by the deviation in A , averaged over 100 Monte Carlo iterations. For convenience of presentation, all three axes are logarithmic. Thus, the coefficient is expressed as the logarithm of the absolute value of the mean. The stochastic coefficient equals the analytical sensitivity coefficient in the cross-hatched area.

that is, from 0.999 to 1.001. Plotted on the vertical axis of Fig. 2 is the logarithm of the absolute value of the mean sensitivity coefficient. Logarithmic scales are therefore used on all axes.

Figure 2 shows that when the error on B is small, the mean stochastic sensitivity coefficient is equal to the analytical sensitivity coefficient (i.e., $\log_{10} 2$) over a wide range of variability for A . As the variability of B increases, the stochastic and deterministic coefficients are equal for much smaller ranges of variability in A . Therefore, as might be expected, the more assumption (2) is violated (i.e. by varying 'another' parameter, B), the more the robustness is lost with respect to assumption (3), i.e. increasing variability in A .

When we consider a simple function like B^4/A^2 , it is relatively easy to find conditions under which the assumptions of sensitivity analysis are satisfied. However, with a complex model, the situation is more difficult. It is simple to say that the assumptions are valid as long as the variabilities are 'small' but difficult to find criteria such as "the assumption is valid up to variabilities of 30%" (Fig. 1). Therefore, it will be necessary to examine a complex non-linear model to gain further insight. The remainder of the paper will be concerned with an analysis of the stream ecosystem model described in the following section.

MODEL AND METHODS

The stream ecosystem model considers mass and nutrient dynamics in a longitudinal stretch of a stream (O'Neill et al., 1979). The state variables of the model are the mass, M , of organic matter, the biomass, H , of heterotrophs, the mass of nitrogen, N , contained in the organic matter and the mean particle size, S , of the organic matter. Particle size is included to account for physical transport and microbial decomposition. The model is based on conservation of mass and considers inputs of organic matter and nitrogen, hydrologic transport of all components and biotic processes such as feeding, respiration, excretion, and predation. The terms of the model are shown in Table I.

It should be remembered that the subject of interest is the comparison of sensitivity analysis with error analysis and not in deriving new insights into the dynamics of stream ecosystems, hence, the choice of the stream ecosystem model is arbitrary. Throughout this presentation, attention is focused on the differences between the two methods of analyses, and not on the implications of the results for stream dynamics.

A detailed description of the model derivation is available elsewhere (O'Neill et al., 1979), so the assumptions of the model are stated only briefly. The physical transport of organic matter mass and nitrogen is assumed to be inversely proportional to the mean diameter of particles. The feeding rate is considered to be a function of the available food and the biomass of consumers. As food becomes very abundant, feeding is a function of consumer biomass. As food becomes very scarce, feeding is a function of the available food supply. Heterotrophic growth rate is limited by food quality (i.e. nutrient concentration) and respiration is assumed to be 50% of assimilated energy. Heterotrophic drift and predation losses are assumed to be density-dependent. Microbial decomposition is modeled as an enzymatic process limited by available surface of the particles.

This model was chosen because it is small, yet mathematically complex. With only four variables and 11 parameters, the task of writing explicit sensitivity equations was feasible and computer costs for multiple comparisons were reasonable. Table I shows that the model contains a number of non-linear terms, including unusual expressions such as the state variable S raised to the $1/3$ or $2/3$ power. As a result of this complexity, higher-order interactions can be expected which provide a good test of the assumptions of sensitivity analysis. In addition, the authors' previous work with the model (O'Neill et al., 1979) provided us with reasonable estimates of the parameter values and allowed us to estimate how well these parameters could be measured in laboratory or field experiments (Table II).

To describe the sensitivity equations, consider a system described by the equations

$$\frac{dx_i}{dt} = f_i(\alpha_1, \alpha_2, \dots, \alpha_m, x_1, x_2, \dots, x_n) \quad (1)$$

TABLE I

The lumped parameter stream model (the first column is the left-hand side of the equations, and the terms on the right-hand side are shown in the subsequent columns)

Parameter	Litter fall	Through-fall and emergence	Physical transport	Heterotrophic feeding	Microbial respiration	Excretion	Assimilation	Predation and emergence
Mass of organic matter (dM/dt)	$10L$		$-0.1aMS^{-1/3}$	$\frac{-2gNH}{C_h(M+H)}$	$-dS^{-2/3}M$			
Heterotrophic biomass (dH/dt)				$\frac{gNH}{C_h(M+H)}$				$-0.1pH^2$
Nitrogen (dN/dt)	$10LC_1$	$10T$	$-0.01bH^2$				$\frac{-gNH}{M+H}$	
Particle size (dS/dt)	$\frac{10LS}{MS_1}(S_1 - S)$		$-0.1aNS^{-1/3}$	$\left(1 - \frac{2N}{C_hM}\right) \frac{gHS}{M+H}$	$-dS^{1/3}$	$-\left(\frac{1-2N}{C_hM}\right) \frac{hgHS}{M+H}$		

TABLE II

A brief description of each parameter in the lumped parameter stream model (the mean values and variance used in the realistic variance case are shown in the latter two columns)

Parameter	Description	Mean	Coefficient of variance
L	Litter fall	1.5	0.50
a	Organic matter discharge coefficient	0.01	0.20
g	Heterotrophic feeding rate with no food limitation	0.3	0.33
C_h	Nitrogen concentration in heterotrophs	0.11	0.10
d	Microbial respiration coefficient	0.002	0.50
b	Heterotrophic drift coefficient	0.001	0.50
p	Heterotrophic predation/emergence coefficient	0.02	0.50
C_1	Nitrogen concentration in litter fall	0.01	0.20
T	Nitrogen input in processes other than litter fall	0.0001	10.00
S_1	Average leaf size	0.05	0.50
h	Number of particles excreted per particle consumed	3.08	30.00

$$x_i(t_0) = x_i^0, i = 1, 2, \dots, n$$

where x_1, x_2, \dots, x_n is the set of state variables, and $\alpha_1, \alpha_2, \dots, \alpha_m$ is the set of parameters. To determine the behavior of this system when the parameters are perturbed, we examine the sensitivity equations

$$\frac{\partial}{\partial \alpha_j} \frac{dx_i}{dt} = \frac{\partial f_i(\alpha_1, \alpha_2, \dots, \alpha_m, x_1, x_2, \dots, x_n)}{\partial \alpha_j} \quad (2)$$

$$\frac{\partial x_i(t_0)}{\partial \alpha_j} = \frac{\partial x_i^0}{\partial \alpha_j} = 0; i = 1, 2, \dots, n; j = 1, 2, \dots, m$$

Equation 2 describes the variation in the state variable, x_i , for an arbitrarily small perturbation of the parameter, α_j (Tomovic and Karplus, 1963). A parameter perturbation has no effect on the initial conditions, so $dx_i(0)/dt = 0$ for all i and j . The partial derivative indicates that parameter j is perturbed while all other parameters are held constant; however, the effect of the perturbation on all state variables must be taken into account. Assuming that the order of differentiation can be reversed and using the chain rule, eq. 2 can be rewritten

$$\frac{d}{dt} \frac{\partial x_i}{\partial \alpha_j} = \frac{\partial f_i}{\partial \alpha_j} + \sum_{k=1}^n \frac{\partial f_i}{\partial x_k} \frac{\partial x_k}{\partial \alpha_j} \quad \frac{\partial x_i(t_0)}{\partial \alpha_j} = 0 \quad (3)$$

$$(i = 1, 2, \dots, n) \quad (j = 1, 2, \dots, m)$$

where the first term on the right-hand side of eq. 3 gives the explicit effect

due to a parameter perturbation, and the second term gives the effect implicit in the variations of the state variables. Now, examining the model given in Table I, there are four state variables and 11 parameters (Table II). Unless the stream is completely uncoupled (only one state variable appearing in each equation), the effects that a variation in one state variable due to a parameter perturbation has on another state variable must be considered. The more closely connected the system, the more complex this relationship will be. Since, with one exception, every state variable appears in every equation (Table I), we would expect the sensitivity relationships for this model to be very complex. Let the symbol $x = y_{a=A}$ indicate that $x = y$ if $a = A$, and that $x = 0$ if $a \neq A$. Then using eq. 3 and Table I, we obtain

$$\begin{aligned} \frac{d}{dt} \frac{\partial M}{\partial \alpha} = & 10 \left| \frac{\partial M}{\partial \alpha} \right|_{\alpha=L} - 0.1MS^{-1/3} \left| \frac{\partial M}{\partial \alpha} \right|_{\alpha=a} - \frac{2NH}{C_h(M+H)} \left| \frac{\partial M}{\partial \alpha} \right|_{\alpha=g} + \frac{2gNH}{C_h^2(M+H)} \left| \frac{\partial M}{\partial \alpha} \right|_{\alpha=C_h} - S^{-2/3}M \left| \frac{\partial M}{\partial \alpha} \right|_{\alpha=d} \\ & - \left(0.1aS^{-1/3} \frac{2gNH}{C_h(M+H)^2} + dS^{-2/3} \right) \frac{\partial M}{\partial \alpha} + (0.1a + 2dS^{-1/3}) \frac{S^{-4/3}M}{3} \frac{\partial S}{\partial \alpha} \\ & - \frac{2gH}{C_h(M+H)} \frac{\partial N}{\partial \alpha} - \frac{2gMN}{C_h(M+H)^2} \frac{\partial H}{\partial \alpha} \end{aligned} \quad (4)$$

$$\begin{aligned} \frac{d}{dt} \frac{\partial H}{\partial \alpha} = & \frac{NH}{C_h(M+H)} \left| \frac{\partial H}{\partial \alpha} \right|_{\alpha=g} - \frac{gNH}{C_h^2(M+H)} \left| \frac{\partial H}{\partial \alpha} \right|_{\alpha=C_h} - 0.01H^2 \left| \frac{\partial H}{\partial \alpha} \right|_{\alpha=b} - 0.1H^2 \left| \frac{\partial H}{\partial \alpha} \right|_{\alpha=p} - \frac{gNH}{C_h(M+H)^2} \frac{\partial M}{\partial \alpha} \\ & + \left(\frac{gMN}{C_h(M+H)^2} - (0.02b + 0.2p)H \right) \frac{\partial H}{\partial \alpha} - \frac{gH}{C_h(M+H)} \frac{\partial N}{\partial \alpha} \end{aligned} \quad (5)$$

$$\begin{aligned} \frac{d}{dt} \frac{\partial N}{\partial \alpha} = & 10L \left| \frac{\partial N}{\partial \alpha} \right|_{\alpha=C_L} + 10C_L \left| \frac{\partial N}{\partial \alpha} \right|_{\alpha=L} + 10 \left| \frac{\partial N}{\partial \alpha} \right|_{\alpha=T} - 0.1NS^{-1/3} \left| \frac{\partial N}{\partial \alpha} \right|_{\alpha=a} \\ & - \frac{NH}{M+H} \left| \frac{\partial N}{\partial \alpha} \right|_{\alpha=g} + \frac{gNH}{(M+H)^2} \frac{\partial M}{\partial \alpha} - \frac{gMN}{(M+N)^2} \frac{\partial H}{\partial \alpha} \\ & - \left(\frac{gH}{M+H} + 0.1aS^{-1/3} \right) \frac{\partial N}{\partial \alpha} + \frac{0.1aNS^{-4/3}}{3} \frac{\partial S}{\partial \alpha} \end{aligned} \quad (6)$$

$$\begin{aligned} \frac{d}{dt} \frac{\partial S}{\partial \alpha} = & \frac{10S(S_1 - S)}{MS_1} \left| \frac{\partial S}{\partial \alpha} \right|_{\alpha=L} + \frac{10S^2L}{MS_1^2} \left| \frac{\partial S}{\partial \alpha} \right|_{\alpha=S_1} - S^{-1/3} \left| \frac{\partial S}{\partial \alpha} \right|_{\alpha=d} \\ & - \frac{2NS(h-1)H}{C_h^2M(M+H)} \left| \frac{\partial S}{\partial \alpha} \right|_{\alpha=C_h} - S \left(1 - \frac{2N}{C_hM} \right) \frac{gH}{M+H} \left| \frac{\partial S}{\partial \alpha} \right|_{\alpha=h} - S \left(1 - \frac{2N}{C_hM} \right)^{(h-1)} \frac{H}{M+H} \left| \frac{\partial S}{\partial \alpha} \right|_{\alpha=g} \end{aligned}$$

$$\begin{aligned}
& - \left[\frac{10L(S_1 - S)}{M^2 S_1} + \left(\frac{2N(h-1)}{C_h M^2} - \left(1 - \frac{2N}{C_h M} \right) \frac{h-1}{M+H} \right) \frac{gH}{M+H} \right] S \frac{\partial M}{\partial \alpha} \\
& - S \left(1 - \frac{2N}{C_h M} \right)^{(h-1)} \frac{gM}{(M+H)^2} \frac{\partial H}{\partial \alpha} + \frac{2S}{C_h M} \frac{(h-1)}{M+H} \frac{gH}{M+H} \frac{\partial N}{\partial \alpha} \\
& + \left[\frac{10L}{MS_1} (S_1 - 2S) - \frac{dS^{-2/3}}{3} - \left(1 - \frac{2N}{C_h M} \right)^{(h-1)} \frac{gH}{M+H} \right] \frac{\partial S}{\partial \alpha} \quad (7)
\end{aligned}$$

Since there are four state variables and 11 parameters, eqs. 4–7 represent 44 equations. These, together with the four state variable equations, give 48 simultaneous differential equations which were solved numerically.

For comparison with the results from the sensitivity equations, Monte Carlo simulations of the model were also performed with each parameter considered as a random variable. Each Monte Carlo iteration was performed by selecting model parameters from independent normal distributions (specified by the means and variances in Table II) and simulating the model for 150 days. When random values less than 0.0 were generated, they were discarded and another random selection was made. Five hundred iterations of the Monte Carlo process were performed for each case considered.

RESULTS

When the coefficient of variation on all parameters is set to 1%, the assumptions of sensitivity analysis are reasonably met since the change in the parameter of interest is small (assumption 3), and other parameters in the model are relatively constant (assumption 2). Table III gives the values of the four state variables at day 150 and their coefficients of variation. It is clear that model results at 1% variability are well behaved. However, the coefficient of variation on H is twice as large as for the other variables. There are two reasons for this increase. First, the model equation for H (Table I) contains a term, H^2 , which we have shown (O'Neill et al., 1980; Gardner et al., 1980b) to cause amplification of parameter error. Second, the magnitude of the derivative of $H(dH/dt)$ can be quite large relative to the size of H (O'Neill et al., 1979). As a result, the values of H can show large deviations from the mean value, resulting in larger variability.

Statistical analyses of the 1% results show the similarity of sensitivity analysis and error analysis when parameter variabilities are small. If one graphs the values of a state variable, x , as a function of one parameter, p (all other parameters assumed to be constant), the sensitivity coefficient represents the slope, S , of a line tangent to the curve, $x = f(p)$, at the mean value of the parameter. The least-squares estimate of the slope of a line, B' , drawn through the Monte Carlo values with p varied around its mean, is the analogous statis-

TABLE III

Means and coefficients of variation (CV) for 500 Monte Carlo iterations of the stream model at day 150 with parameters varied at 1% of their mean value and at realistic variance levels (RVL) (See Table I)

State variable	1% Case		RVL Case	
	Mean	CV	Mean	CV
<i>M</i>	1634.5	1.1	290.9	218.5
<i>H</i>	14.5	2.7	27.8	66.9
<i>N</i>	19.7	1.3	5.8	165.3
<i>S</i>	0.3	1.7	0.04	296.8

tical measure. When variability of p is small, the scatter of points produced by iterative solutions of the model approximates a straight line and B' is similar to S .

Figure 3 shows the relationship between S and B' at day 150. The values of B' are the partial regression coefficients obtained by multiple regression analysis. The B' values represent the relationship between x and p with the effect of the other parameters removed (Draper and Smith, 1966), i.e. the least-squares estimates of the slope of a line through the residuals which remain after the effects of the variance of all other parameters have been removed.

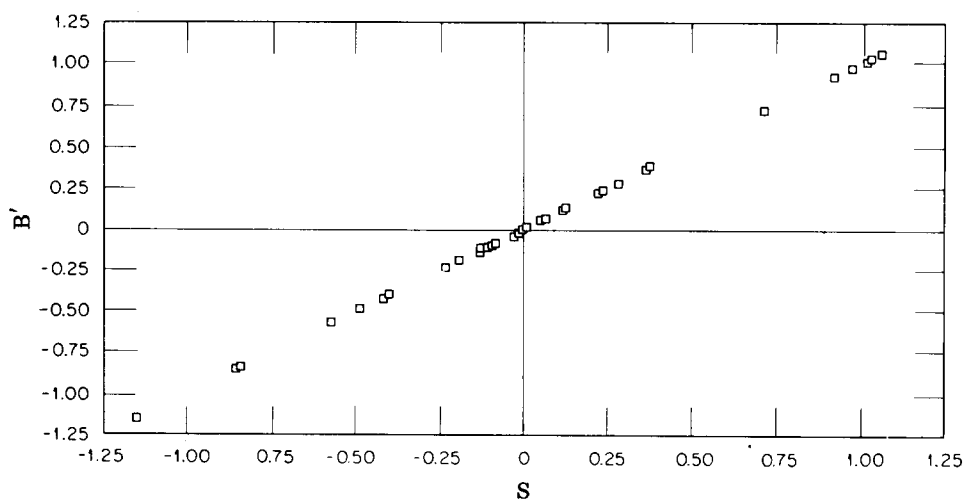


Fig. 3. Relationship between B' and S at day 150 of the stream model. B' is the partial regression coefficient of each parameter on each state variable obtained from 500 Monte Carlo iterations with parameters varied at 1% of their mean value. S is the sensitivity coefficient normalized by multiplying by p/x , where p and x are the parameter and state value, respectively.

TABLE IV

Rank ordering of the sensitivity and B' coefficients, squared partial correlation (r_p^2), and squared simple correlation (r_s^2) for the 1% variance case. Each entry indicates the state vs. parameter comparison involved (see Tables I and II)

Rank	Sensitivity	B'	r_p^2	r_s^2
1	$M-d$	$M-d$	$M-L$	$M-L$
2	$M-a$	$M-a$	$N-L$	$N-L$
3	$M-C_1$	$M-T$	$S-S_1$	$S-S_1$
4	$M-T$	$M-C_1$	$N-C_1$	$H-C_h$
5	$M-p$	$M-p$	$M-d$	$H-g$
6	$H-d$	$H-d$	$H-C_h$	$S-g$
7	$M-C_h$	$M-C_h$	$S-g$	$N-C_1$
8	$N-C_1$	$N-C_1$	$H-g$	$H-C_1$
9	$H-C_1$	$H-C_1$	$H-C_1$	$H-p$
10	$M-L$	$M-b$	$M-S_1$	$M-d$
SRCC *		0.997	0.801	0.677
Max % **			100.0	76.7
Min % **			99.9	11.4

* SRCC is the Spearman rank correlation coefficient.

** Max % and Min % are the maximum and minimum percent of the variance accounted for by the first and last correlation listed.

The values of B' and S in Fig. 3 are essentially identical, and the correlation between them is 0.999. The ranking of the model parameters from most important to least important based on the two parameters is essentially the same (Spearman rank correlation coefficient of 0.997, Table IV). No additional information has been gained by numerous iterations of the model to obtain an estimate of B' . For this comparison, the sensitivity results suffice.

The values of B' and S indicate the direction and magnitude of change in x given a change in p ; therefore, neither relates directly to prediction uncertainty. The statistic of immediate interest is the ratio of the variance of x to the variance of p , i.e. the squared correlation coefficient R^2 (Harris, 1975), since this represents the improvement in $f(p)$ that can be expected from a better measurement of p .

A comparison of the ranking of parameters (Table IV) shows that even though 1% variability in parameters is relatively small, and the assumptions of sensitivity analysis seem to be met, the results of sensitivity analysis can be misleading. The variable/parameter pairs that show the highest sensitivities are quite different from those that show the highest squared correlations (Table IV). Thus, improved measurements of litterfall, L (highest correlation coefficients), will improve the estimates of M and N predicted from the model. Improved measurements of parameters d or a (highest sensitivities) result in little or no reduction in the variance of model output.

Two correlation coefficients are available for each parameter with respect to each state variable. The partial correlation coefficient, r_p , when squared, represents the percentage of the variability in x accounted for by variability in p , when the linear effects of all other parameters have been removed. This estimates the improvement in model predictions which would result from reducing the variability in p , given that the variability of the other parameters can be controlled. The second coefficient, the simple correlation coefficient, r_s , when squared, represents the percentage of the variability in x accounted for by variability in p , when the variability of the other parameters is uncontrolled.

If $f(p)$ is a linear function of p , r_p will be approximately unity. Once the effects of the other parameters have been removed, the residuals will lie around a straight line, described by the slope B' . The percentage of the variability in $f(p)$ described by this relationship will therefore approach 100%.

When the relationship between p and $f(p)$ is non-linear, the residuals cannot be exactly described by a straight line and B' cannot explain all of the changes in $f(p)$. As a result, r_p will be less than 1.0, indicating that higher-order terms (i.e. non-linear in the statistical sense) are required to explain the relationship. Thus, we can define

$$C = 1 - R^2, \quad (8)$$

as an indicator of the importance of higher-order effects in the relationship between a parameter and state variable.

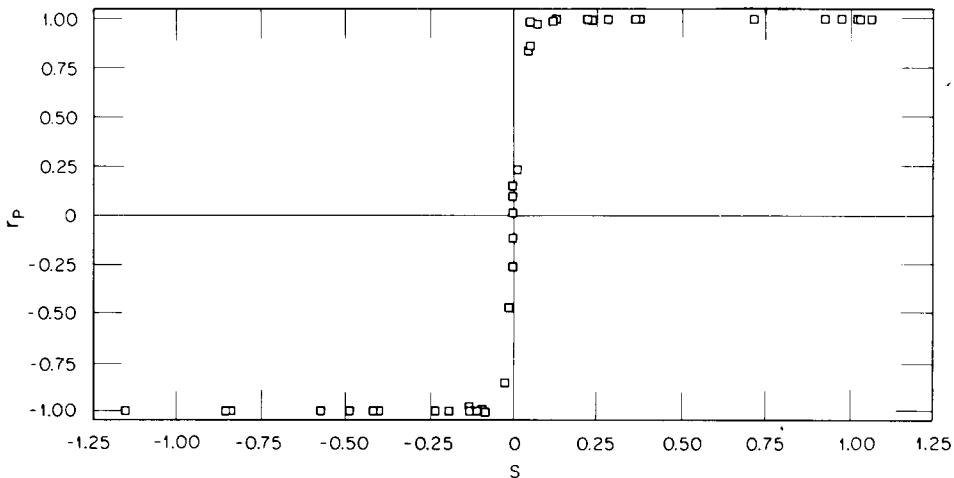


Fig. 4. Relationship between r_p^2 and S at day 150 of the stream model. r_p^2 is the partial correlation coefficient of each parameter on each state variable from 500 Monte Carlo iterations of the stream model with parameters varied at 1% of their mean value. S is the sensitivity coefficient normalized by multiplying by p/x , where p and x are the parameter and state value, respectively.

Figure 4 shows the relationship between the partial correlation coefficient, r_p , and the sensitivity coefficient, S for the 1% variability case. When the sensitivities are large, the correlation coefficients are approximately 1.0. When the sensitivities are near zero, higher-order effects are evident and the r_p values are significantly different from 1.0 or -1.0. When sensitivities are large, a straight line drawn through the Monte Carlo results accounts for a large percentage of the variability in $f(p)$. If the sensitivity coefficient is near zero, the relationship between $f(p)$ and p is approximately a horizontal straight line. When the slope is zero, the amount of variance in x explained by the variance in p declines sharply, and the remaining deviations from the horizontal line must be accounted for by higher-order effects.

The importance of higher-order effects becomes more evident when all parameters are varied by more than 1% of their means. As the variability in p is increased, there is a decreasing likelihood that $f(p)$ can be adequately approximated by a straight line. The smaller the percentage of the variability in $f(p)$ that can be explained by a straight line, the less important is the slope of the tangent (i.e. the sensitivity coefficient) as an indicator of the relative importance of parameters.

The differences between sensitivity analysis and error analysis become most evident when we allow the variability of individual parameters to take on 'realistic' values. In Table II we have estimated each coefficient of variation based on our intuition of how well each parameter could be measured in the field or laboratory. The values we have chosen are certainly open to debate. The important point is that parameters differ in their associated error terms and that at least some of the parameters have large variances associated with them.

Table III gives the means and coefficients of variation for the four state variables on day 150. For the realistic variance levels (RVL) case (Table II), the mean values are significantly different from the 1% case and the variability is greatly increased. The expected value of the model with parameters varied realistically is not equal to the model output with each parameter taken at its mean value (approximately equal to the 1% case). Therefore, assumption (1) of sensitivity analysis is clearly violated. The bias is particularly evident in the M and S variables which are reduced by an order of magnitude. Perhaps more striking is the fact that variable H now has the smallest, instead of the largest, coefficient of variation.

It is also clear that higher-order effects now dominate the analysis. Figure 5 shows that the majority of variable-parameter pairs in the 1% case had values of C (eq. 8) of 0–20%. That is, higher-order effects, violating assumptions (2) and (3), were not important. In the RVL case, most correlations are dominated by higher-order effects with C values between 80 and 100%. It is apparent that large variances, characteristic of ecological problems, can seriously violate the assumptions of sensitivity analysis.

Table V shows the rank ordering of the variable-parameter pairs according to sensitivity coefficients, B' , and both simple and partial correlations for the

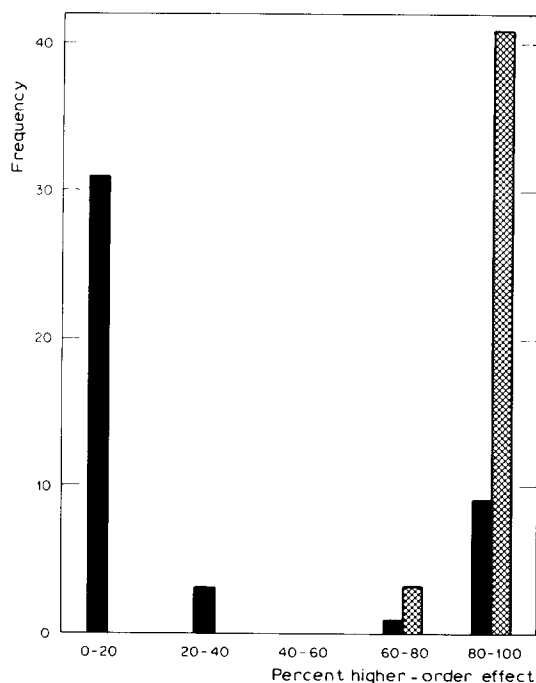


Fig. 5. Histogram showing the frequency of higher order effects occurring in the relationship between state values and parameters. The percent higher order effect is estimated by $(1 - r_p^2) \times 100$, where r_p^2 is the partial correlation between parameters and state values. The solid bars are the 1% case and the crosshatched bars are the RLV case (see Table II).

RVL. Since sensitivity and B' estimate the same linear slope, rankings according to these two coefficients are similar although the Spearman rank correlation coefficient has dropped from 0.997 (Table IV) to 0.716. The ranking according to regression coefficients, however, is quite different. The only parameter in common is litterfall, L . Therefore, decisions on which parameters should be measured more accurately would be radically different depending on the technique applied.

To illustrate the difference between decisions based on sensitivity analysis and error analysis, we simulated the situation in which the field researcher received feedback from each technique. Thus, we took the four parameters showing the highest sensitivities, i.e. d , a , C_1 , T (Table V) for one case and the four parameters showing the largest values of r_s^2 , i.e. p , g , h , and L for the other. In both cases, we reduced the variance of these parameters by a factor of two and ran 500 iterations.

Halving the variance of parameters with the highest sensitivities reduced the coefficient of variation on state variables M , N and S by only 2–6%. The coefficient of variation on H was actually increased by 22%. In contrast,

TABLE V

Rank ordering of the sensitivity and B' coefficients, squared partial correlation (r_p^2), and squared simple correlation (r_s^2) for the realistic variance case. Each entry indicates the state vs. parameter comparison involved (see Tables I and II)

Rank	Sensitivity	B'	r_p^2	r_s^2
1	$M-d$	$M-d$	$N-g$	$H-p$
2	$M-a$	$M-b$	$H-p$	$N-g$
3	$M-C_1$	$M-C_1$	$N-L$	$M-h$
4	$M-T$	$M-a$	$M-g$	$M-g$
5	$M-p$	$M-p$	$M-h$	$N-L$
6	$H-d$	$M-g$	$N-h$	$N-h$
7	$M-C_h$	$M-C_h$	$H-L$	$S-h$
8	$N-C_1$	$H-T$	$M-L$	$H-L$
9	$H-C_1$	$H-C_1$	$S-h$	$S-g$
10	$H-L$	$H-b$	$S-g$	$M-L$
SPCC *		0.716	0.658	0.68
Max % **			35.5	29.4
Min % **			10.3	6.6

* SPCC is the Spearman rank correlation coefficient.

** Max % and Min % are the maximum and minimum percent of the variance accounted for by the first and last correlation listed.

halving the variability of the four parameters with the largest simple correlation coefficient reduced the coefficients of variation on the four state variable between 17 and 31%. Thus, a far more significant reduction in prediction uncertainty resulted from ranking the variables according to the correlation coefficient and utilizing this ranking as a basis for deciding which parameters should be measured more accurately.

DISCUSSION

This study of the stream ecosystem model indicates that sensitivity analysis has limited applicability for modeler-researcher interactions in ecology. A sensitivity coefficient estimates the linear relationship between a parameter and model output. However, this relationship is not of primary importance to the researcher interested in reducing the uncertainty associated with model predictions.

In many scientific problems, model parameters can be measured with great accuracy. In these cases, the results of sensitivity analysis are directly applicable. However, in ecological applications, large variance on measurements is the rule and sensitivity analysis is less useful in this situation.

Even with small variabilities (Table IV), the variance in model output due to each parameter is not adequately estimated by sensitivity. Since the imme-

diated concern is to reduce prediction uncertainty through allocation of research effort, the relevant information concerns which parameters, if measured with smaller variances, will result in the greatest reduction in prediction uncertainty. This information is provided by the correlation coefficients which rank order the model parameters quite differently from sensitivity (Table V).

Our results indicate that basing research effort on the correlation coefficients results in greater reduction in prediction uncertainty. In addition, our analysis of higher order effects (Fig. 5) shows that the underlying assumptions of sensitivity analysis are seriously violated when realistic variances on model parameters are considered.

If the modeler is interested in the underlying error propagation properties of a model, the partial correlation coefficient will be of greatest interest. This coefficient estimates the influence on prediction uncertainty due to the variability of one parameter after the linear effects of variability in other parameters have been removed. Therefore, it provides a tool for examining the isolated effects of each parameter.

However, in providing information to the researcher, the simple correlation coefficient may be more relevant. In the field situation all components of the system are subject to variability and are measured with error. These other sources of variability cannot be controlled, and their influence on prediction uncertainty cannot be removed as in partial correlation analysis. In this situation, the ranking of parameters according to the simple correlation coefficient is most relevant.

The stream ecosystem model analyzed in this study is non-linear, and the statistical analyses we have applied assume a linear relationship between parameters and model output. The importance of higher-order effects, particularly with realistic variances on parameters (Fig. 5), clearly shows that a linear analysis is only an approximation. Residual effects due to non-linear relationships and interaction terms have not been considered in the analysis.

Therefore, at present, we can only suggest that correlation coefficients calculated from Monte Carlo results are the most relevant approach for ranking model parameters. Further development is required in order to apply non-linear statistics which will account for all, or most of the higher-order effects.

REFERENCES

- Argentesi, F. and Olivi, L., 1976. Statistical sensitivity analyses of a simulation model for the biomass-nutrient dynamics in aquatic ecosystems. In: Proc. Summer Computer Simulation Conf., Simulation Councils, La Jolla, CA, pp. 389-393.
- Beck, M.B., 1979. Model structure identification from experimental data. In: E. Halfon (Editor), *Theoretical Systems Ecology*. Academic Press, New York, NY, pp. 259-289.
- Beck, M.B., 1981. *Uncertainty Analysis of Aquatic Ecosystem Models*, Int. Inst. Appl. Syst. Anal., Vienna (in press).
- Draper, N.R. and Smith, H., 1966. *Applied Regression Analysis*. Wiley, New York, NY, 407 pp.
- Gardner, R.H. and O'Neill, R.V., 1981. Parameter uncertainty and model predictions. In: M.B. Beck (Editor), *Uncertainty Analysis of Aquatic Ecosystem Models*. Int. Inst. Appl. Syst. Anal., Vienna (in press).

- Gardner, R.H., Huff, D.D., O'Neill, R.V., Mankin, J.B., Carney, J. and Jones, J., 1980a. Application of error analysis to a marsh hydrology model. *Water Resour. Res.* 16: 659—664.
- Gardner, R.H., O'Neill, R.V., Mankin, J.B., Carney, J.H. and Kumar, D., 1980b. Comparative error analysis of six predator-prey models. *Ecology* 61: 323—332.
- Halfon, E., 1975. The systems identification problem and development of ecosystem models. *Simulation*, 38: 149—152.
- Harris, R.J., 1975. *A Primer of Multivariate Statistics*. Academic Press, New York, NY, 332 pp.
- Innis, G.S. (Editor), 1978. *Grassland Simulation Model*. Springer-Verlag, New York, NY, 298 pp.
- Jørgensen, S.E. (Editor), 1979. State-of-the-art in Ecological Modelling. *Proc. Conf. Ecol. Modelling*, Copenhagen, 28 August—2 September 1978. *Int. Soc. Ecol. Modelling*, Copenhagen, 891 pp.
- O'Neill, R.V., 1971. Tracer kinetics in total ecosystems: a systems analysis approach. In: *Nuclear Techniques in Environmental Pollution*. *Int. At. Ener. Comm.*, Vienna, pp. 693—705.
- O'Neill, R.V., 1973. Error analysis of ecological models. In: D.J. Nelson (Editor), *Radio-nuclides in Ecosystems*. U.S. At. Energ. Comm., Washington, DC, pp. 898—908.
- O'Neill, R.V. and Gardner, R.H., 1979. Sources of uncertainty in ecological models. In: B.P. Zeigler, M.S. Elzas, G.J. Klir and T.I. Oren (Editors), *Methodology in Systems Modelling and Simulation*. North-Holland Press, Amsterdam, pp. 447—463.
- O'Neill, R.V., Elwood, J.W. and Hildebrand, S.G., 1979. Theoretical implication of spatial heterogeneity in stream ecosystems. In: G.S. Innis and R.V. O'Neill (Editors), *Systems Analysis of Ecosystems*. International Cooperative Publishing House, Fairland, MD, pp. 79—101.
- O'Neill, R.V., Gardner, R.H. and Mankin, J.B., 1980. Analysis of parameter error in a nonlinear model. *Ecol. Modelling*, 8: 297—311.
- Shugart, H.H., Goldstein, R.A., O'Neill, R.V. and Mankin, J.B., 1974. TEEM: a terrestrial ecosystem energy model for forests. *Oecol. Plant.*, 9: 231—264.
- Steinhorst, R.K., Hunt, H.W., Innis, G.S. and Haydock, K.P., 1978. Sensitivity analysis of the ELM model. In: G.S. Innis (Editor), *Grassland Simulation Model*. Springer-Verlag, New York, NY, pp. 231—256.
- Tomovic, R. and Karplus, W.J., 1963. *Sensitivity analysis of dynamic systems*. McGraw-Hill, New York, NY, 394 pp.
- Van Winkle, W. (Editor), 1977. *Assessing the Effects of Power-Plant Induced Mortality on Fish Populations*. Pergamon, New York, NY, 380 pp.